

PROGRAM DESCRIPTION

Centre d'Etude du Polymorphisme Humain (CEPH): Collaborative Genetic Mapping of the Human Genome

JEAN DAUSSET,* HOWARD CANN,*
DANIEL COHEN,* MARK LATHROP,*
JEAN-MARC LALOUEL,† AND RAY WHITE†¹

*Centre d'Etude du Polymorphisme Humain (CEPH), 27 rue Juliette Dodu, 75010 Paris, France; and †Howard Hughes Medical Institute and Department of Human Genetics, University of Utah Health Sciences Center, Salt Lake City, Utah 84132

Elsewhere in this issue of *Genomics* is the first consortium map, that of chromosome 10, from the CEPH collaboration to map the human genome genetically. The map is truly a collaborative achievement, in that the underlying genotypes represent the efforts of laboratories collaborating with each other and with CEPH to produce a primary genetic map of the genome, consisting of polymorphic markers placed at approximately 20-cM intervals along each of the human autosomes and the X chromosome. Such a map provides a tool for the systematic localization of genes that determine inherited diseases and of other genes of interest. Genetic localization can be the first step in the development of diagnostic tests and isolation of a disease-determining gene. The purpose of this program description is to provide information about CEPH and the basis and nature of the collaboration.

CEPH

The Centre d'Etude du Polymorphisme Humain (CEPH)² is a nonprofit research institute that makes available to the scientific community a valuable research resource. CEPH is committed to (1) make available to the scientific community DNA samples from a panel of reference families for the determination of genotypes for various DNA polymorphisms which may be used for the construction of the genetic map of the human genome and for other research areas dependent on access to such a common set of families, and (2) provide to the contributors of genotypes a compilation of all data that accumulate on the panel of families.

CEPH Collaboration

The CEPH collaboration to map the human genome was organized in 1984 to hasten construction of a primary human

genetic map with DNA polymorphisms (Botstein *et al.*, 1980). A key premise of the CEPH collaboration is that the human genetic map will be efficiently achieved by collaborative research on DNA from the same sample of families. To this end, CEPH provides to collaborating investigators high-quality cellular DNA produced from cultured lymphoblastoid cell lines (LCL) derived from each member of a reference panel of large nuclear families/pedigrees and a database contributed to and shared by these investigators. Collaborating investigators determine genotypes with their probes and the DNA from the CEPH panel to test the families for segregation of these genetic markers. They then contribute the genotypes to CEPH for preparation of a database which is returned to them for linkage analysis and map construction. As of October 1, 1989, 63 research laboratories in the United States (36), Canada (2), Europe (20), South Africa (2), Japan (2), and Australia (1) collaborate with CEPH in this manner.

CEPH Reference Family Panel

Families with large sibships, living parents, and grandparents are especially informative for linkage mapping (White *et al.*, 1986). From 100 families available from various sources, selected not for disease but for large sibship size, an initial group of 40 families was defined for the CEPH reference panel by the original group of collaborating investigators. Table 1 shows the geographic origins of these families and the contributors of the LCLs to CEPH. These are Caucasian families. The mean sibship size for these 40 families, based on those individuals for whom there are LCLs, is 8.3; no family has less than 6 offspring, and 23 families have 8 or more offspring. LCLs are available for all 4 grandparents in each of 29 families of the reference panel.

LCLs of the reference panel are stored in liquid nitrogen at three geographically separate repositories: Paris and Lyon, France; and Salt Lake City, Utah. These LCLs are not distributed by CEPH within or outside of the collaboration. They are used only as a source of DNA for the collaboration. LCLs for 11 of the panel families are available from the NIGMS Human Genetic Mutant Cell Repository located at the Coriell Institute for Medical Research in Camden, New Jersey.

Approximately 20 mg of DNA is prepared, as needed, from each LCL, and aliquots of 200 or 400 μ g are distributed to each collaborating investigator. There is no charge to collaborating investigators for the DNA. The DNA is prepared by classical methods, scaled up to preparation of milligram amounts, which include lysis with proteinase K and SDS, extraction with phenol, and precipitation with isopropanol. The quality of each preparation of CEPH DNA is routinely controlled by testing for concentration, molecular size, digestibility with two different restriction enzymes, and contamination with vector sequences (Southern blot hybridized with a cosmid vector). In addition, each DNA preparation is hybridized with at least two highly variable ("minisatellite") probes (Wong *et al.*, 1987) to confirm identity of the LCL source and to detect contamination with DNA from another individual.

¹ All authors are members of the CEPH Executive Committee.

² Founded in 1983 by J.D. and D.C.

SPECIAL FEATURE

TABLE 1
Sources of CEPH Family Panel

Utah	27 families*	R. White
Franco	10 families	J. Dausset
Venezuela	2 families*	J. Gusella
Pennsylvania (Old Order Amish)	1 family*	J. Egeland

* LCLs from 9 families available from NIGMS Human Genetic Mutant Cell Repository, Camden, NJ.

† LCLs from 1 of 2 families available from NIGMS Human Genetic Mutant Cell Repository.

‡ LCLs available from NIGMS Human Genetic Mutant Cell Repository.

DNA Polymorphisms

The family panel DNA is being tested for at least 1200 DNA polymorphisms. The probes being used within the collaboration detect polymorphism within restriction enzyme sites of low copy number or unique DNA segments, due to varying numbers of tandemly repeated, relatively small sequences, or VNTRs (Nakamura *et al.*, 1987), and of centromere-associated alpha satellite DNA (Willard *et al.*, 1986). Some collaborating investigators have begun to use genetic markers with family panel DNA that are based on detection of length variations of very small sequences amplified by PCR (Weber and May, 1989; Litt and Luty, 1989). Phenotypes for classical, serological, and electrophoretic polymorphic markers are also available for the family panel.

CEPH Database

CEPH collaborating investigators have agreed on two basic rules concerning their participation within the collaboration: (1) A CEPH collaborating investigator is committed to screening the reference panel by determining genotypes for all 40 parent pairs with each probe and enzyme combination used and then following segregation in each informative family (at least one heterozygous parent) in his or her own laboratory or collaboratively. (2) Collaborating investigators are committed to sending genotypic data to CEPH for inclusion in the CEPH database no later than publication of the data themselves or of an article based on the data.

Genotypic data generated from family panel DNA are processed into uniform format by a set of programs developed at CEPH (J.-M.L.) for use with IBM PCs or compatible models and sent to all collaborating investigators. The data files thus prepared are sent to CEPH to be merged into the CEPH database.

There are two components of the database: The CEPH collaborative database is available only within the collaboration. Two classifications of data are recognized in the collaborative database, *unpublished* and *published*. The first are those data that have not been used for a publication. These are privileged data, requiring permission from the contributor for use (e.g., publication) by another collaborating investi-

gator. Published data are automatically available for inclusion in CEPH consortium maps (see below) and after a lag period will be released to the CEPH public database. After 2 years in the collaborative database, unpublished data become published data. The CEPH public database, currently being organized, will contain published data released from the collaborative database. Data in the public database will be available to the general scientific community.

Table 2 summarizes the contents of the CEPH collaborative database. As of July 1989, the database contains genotypic data for 1061 genetic markers localized to all the autosomes and the X chromosome (including the pseudoautosomal region). Approximately 20% of these marker systems have four or more alleles.

New CEPH Activities

As the primary map of the human genome nears completion, there is growing interest in a higher resolution genetic map, perhaps of the order of 1–2 cM. Availability of a higher resolution map will increase the efficiency and precision of localization of genes. In order to support construction of a high-resolution map of the genome, CEPH is in the process of increasing the family panel to 61 large nuclear families/pedigrees. LCLs from 21 additional families have already been received from Utah, and stocks of DNA are being prepared from them. The mean sibship size for the total of 61 families in the augmented panel is 8.5, and LCLs for all 4 grandparents are available for each of 44 families. These 21 families provide an additional advantage for genetic mapping in that they have already been genotyped for approximately 500 genetic markers used in the CEPH collaboration.

CEPH has undertaken a project to enhance the use of the primary map of the genome in localizing genes of interest. The idea is to collect and produce quantities of ready-to-label probes for the mapped primary markers and distribute these to biomedical scientists who wish to localize disease-determining genes and other genes of interest. Probes to be distributed will be those for markers chosen from CEPH consortium maps based on criteria that include heterozygote frequency and position on the genetic map of a chromosome. Collections of probes ("kits") will be available for each chromosome. By using the probes in a kit to test a large kindred or group of families in which a genetic disease with the appropriate chromosome assignment is segregating, an inves-

TABLE 2
CEPH Database

Version	No. of contributing labs	No. of markers	No. of markers with 4 or more alleles	Date
V1	17	171	28	April 1987
V2	29	744	143	March 1988
V3	33	1061	204	July 1989

tigator should be able to find linkage to one, two, and perhaps more markers if a sufficient number of informative meioses are available. Again, if data from a sufficient number of meioses are available, we would expect that a gene of interest without a chromosome assignment could be localized with markers from all of the kits. Investigators using these kits for primary mapped markers will be invited to contribute the genotypes they determine from families being tested with the probes to a database provided by CEPH. This project, being carried out in collaboration with the American Type Culture Collection, is sponsored by the National Institutes of Health.

Progress toward the Primary Genetic Map of the Human Genome

The number of markers for which there are genotypic data in the CEPH database suggests that the primary human genetic map is nearing completion. Partial or nearly complete³ primary linkage maps based on genotypes determined from reference panel DNA with probes from a single or small group of CEPH collaborating laboratories have been published for many of the chromosomes (see, for example, Donis-Keller *et al.*, 1987; O'Connell *et al.*, 1989; Lathrop *et al.*, 1988; Nakamura *et al.*, 1988; Warren *et al.*, 1989). CEPH consortium maps, the first of which appears in this issue of *Genomics*, will provide reference primary linkage maps of each chromosome based on the combined genotypic data within the collaboration.

CEPH Consortium Maps

The primary genetic map of the human genome, based on all genotypes determined from the CEPH families, will be communicated to the scientific community through a series of consortium maps, one for each chromosome. These maps will represent the ultimate validation of genotypic data in the CEPH database. Each consortium map will be prepared by a committee of collaborating investigators who have contributed genotypic data for markers localized to the particular chromosome. A chromosome-specific database, containing genotypes for all relevant markers, will be sent to members of the consortium map committee for the construction of genetic maps. These maps will be circulated within the committee for study and comparison. The committee will meet for the final analysis of these maps, which will be used in the preparation of the consortium map. The consortium map and a report will be published in *Genomics*. It is this process that led to the CEPH consortium map of chromosome 10 presented in this issue of *Genomics*.

As each consortium map is published, the underlying genotypic data will be released to the scientific community. The consortium maps will provide the basis for choosing the primary mapped markers for which kits of probes will be distributed by CEPH to the scientific community in order to

enhance the use of the genetic map for localization of genes that determine disease and other genes of interest.

REFERENCES

1. BOTSTEIN, D., WHITE, R. L., SKOLNICK, M., AND DAVIS, R. W. (1980). Construction of a genetic linkage map using restriction fragment length polymorphisms. *Amer. J. Hum. Genet.* 32: 314-331.
2. DONIS-KELLER, H., GREEN, P., HELMS, C., CARTINHO, S., WEIFFENBACH, B., STEPHENS, K., KEITH, T., BOWDEN, D., SMITH, D., LANDER, E., BOTSTEIN, D., AKOTS, G., REDIKER, K., GRAVIUS, T., BROWN, V., RISING, M., PARKER, C., POWERS, J., WATT, D., KAUFFMANN, E., BRICKER, A., PHIPPS, P., MULLER-KAHLE, H., FULTON, T., NG, S., SCHUMM, J., BRAMAN, J., KNOWLTON, R., BARKER, D., CROOKS, S., LINCOLN, S., DALY, M., AND ABRAHAMSON, J. (1987). A genetic linkage map of the human genome. *Cell* 51: 319-337.
3. LATHROP, M., NAKAMURA, Y., O'CONNELL, P., LEPPERT, M., WOODWARD, S., LALOUEL, J.-M., AND WHITE, R. (1988). A mapped set of genetic markers for human chromosome 9. *Genomics* 3: 361-366.
4. LITT, M., and LUTY, J. A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Amer. J. Hum. Genet.* 44: 397-401.
5. NAKAMURA, Y., LEPPERT, M., O'CONNELL, P., WOLFF, R., HOLM, T., CULVER, M., MARTIN, C., FUJIMOTO, E., HOFF, M., KUMLIN, E., AND WHITE, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 235: 1616-1622.
6. NAKAMURA, Y., LATHROP, M., O'CONNELL, P., LEPPERT, M., BARKER, D., WRIGHT, E., SKOLNICK, M., KONDOLEON, S., LITT, M., LALOUEL, J.-M., AND WHITE, R. (1988). A mapped set of DNA markers for human chromosome 17. *Genomics* 2: 302-309.
7. O'CONNELL, P., LATHROP, G. M., NAKAMURA, Y., LEPPERT, M. L., ARDINORR, R. H., MURRAY, J. L., LALOUEL, J.-M., AND WHITE, R. (1989). Twenty-eight loci form a continuous linkage map of markers for human chromosome 1. *Genomics* 4: 12-20.
8. WARREN, A. C., SLAUGENHAUPT, S. A., LEWIS, J. G., CHAKRAVARTI, A., AND ANTONARAKIS, S. E. (1989). A genetic linkage map of 17 markers on human chromosome 21. *Genomics* 4: 579-591.
9. WEBER, J. L., AND MAY, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Amer. J. Hum. Genet.* 44: 388-396.
10. WHITE, R., LEPPERT, M., BISHOP, T., BARKER, D., BERKOWITZ, J., BROWN, C., CALLAHAN, P., HOLM, T., AND JEROMINSKI, L. (1986). Construction of linkage maps with DNA markers for human chromosomes. *Nature (London)* 319: 101-105.
11. WILLARD, H. F., WAYE, J. S., SKOLNICK, M. H., SCHWARTZ, C. E., POWERS, V. E., AND ENGLAND, S. B. (1986). Detection of restriction fragment length polymorphisms at the centromeres of human chromosomes by using chromosome-specific satellite DNA probes: Implications for development of centromere-based genetic linkage maps. *Proc. Natl. Acad. Sci. USA* 83: 5611-5615.
12. WONG, Z., WILSON, V., PATEL, I., POVEY, S., AND JEFFREYS, A. J. (1987). Characterization of a panel of highly variable minisatellites cloned from human DNA. *Ann. Hum. Genet.* 51: 269-288.

³ A primary linkage map of a chromosome cannot be considered complete until it contains markers for, or quite close to, the telomeres.